

Pandas Exercise 2

Time to practice managing NaN and NaT values.

1. Import the csv "OrangeJuiceDirty.csv" into an ipynb file. OrangeJuiceDirty.csv can be found in this chapter's dropbox. That file contains real prices for Orange Juice CFDs. Some of the datetime and price values have been replaced by text.
2. Convert the first column to datetime.
3. Convert the rest of the columns to numbers.
4. Merge all converted columns to a new dataframe. Let's call this dataframeX.
5. Clean dataframeX in 3 separate ways and output each result in a new dataframe.
 - a. Delete all rows with at least one NaN/NaT value.
 - b. Replace the NaN/NaT values with the average value of its column.
 - c. Replace the NaN/NaT values with the value above it.

Notes:

- For 5b), the NaT values will remain as NaT. Don't fret about that.
- 5c) is more difficult than the rest. It can't be solved by thinking. I strongly recommend you Google the solution. Take this chance to practice your solution Googling skills. If you need help, we have included hints on the following page.
- These aren't the correct ways to clean datetime data. We will talk about doing it right in a later chapter. For now, we just want to focus on the basics of managing NaN/NaT.
- Solutions are found in the file "Dataframes Exercise 2 – Solutions.ipynb". Link: https://www.dropbox.com/sh/4xmb3f0alqrx0mc/AAC2w9_qtGIfcF3_8kuptLQGa?dl=0

Hints:

- For 5c), Google search something along the lines of “pandas dataframe how to fillna with the value above”. The solution should be in one of the top 5 search results.